1985

The Estimation of Probability Density Functions and its Applications in Econometrics

Aman Ullah

Radhey S. Singh

ISSN: 0831-4454
ISBN: 0-7714-0702-5

THE CENTRE FOR DECISION SCIENCES AND ECONOMETRICS

The Estimation of Probability Density Functions and its Applications in Econometrics

Aman Ullah and Radhey S. Singh

TECHNICAL REPORT NO. 6
AUGUST 1985

Centre For Decision Sciences And Econometrics
Social Science Centre
The University of Western Ontario
London, Ontario N6A 5C2

THE ESTIMATION OF PROBABILITY DENSITY FUNCTIONS AND ITS APPLICATIONS IN ECONOMETRICS

Aman Ullah

Department of Economics
University of Western Ontario
London, Ontario, Canada

Radhey S. Singh

Department of Mathematics and Statistics
University of Guelph
Guelph, Ontario, Canada

SUMMARY

In this paper we present a class of nonparametric estimates of densities which are asymptotically unbiased and consistent. We point out various applications of these density estimates in econometrics. Some illustrative examples, using economic data, are also given.

1. INTRODUCTION

It is now well known that parametric inference in econometrics is carried on under various pretensions. For example, consider y and x as two economic variables, say, consumption and income, respectively. Then, first, it is usually assumed that y is stochastic but x is controlled (non-stochastic) when in fact both are stochastic. Second, even if both y and x are stochastic the conditional model ($E(y|_x)$) is used under the assumption that the parameters of the conditional and marginal distributions are variation free so that x is weakly exogenous (Engle et al (1983)). Third, the functional form of the conditional model is taken as linear when in fact it may be nonlinear. Fourth, the joint density (data generating process) of y and x is usually assumed to be normal. These are some basic assumptions, among others, which make the existing empirical research in econometrics look dubious.

Here we explore an alternative procedure which is free of the assumptions indicated above. This alternative is based on the estimation of a probability density function and/or its derivatives and this has drawn considerable attention in the statistical literature, see, e.g., Rosenblatt (1956), Parzen (1962), Cacoullos (1966) and Singh (1978, 1981) among others. However, very few attempts have been made to explore the application of density estimates to any area of applied research although Singh (1977) has pointed out some applications to statistical problems. The modest aim of this paper is to briefly review the literature on density function estimation and investigate its applications in various areas of econometrics.

In Section 2 we present a class of nonparametric estimates of univariate and multivariate densities which are asymptotically unbiased and consistent. Then, in Section 3, we point out the applications of density estimates to several problems that arise in econometrics. Some illustrative examples are presented in Section 4.

2. ESTIMATORS OF A PROBABILITY DENSITY FUNCTION

2.1 Univariate Case

The density function of a random variable X at a point x is defined by:

$$f = f(x) = \frac{d}{dx} F(x) = \lim_{h \to o} \frac{F(x+h/2)-F(x-h/2)}{h} \tag{2.1}$$

$$= \lim_{h \to o} h^{-1} P[x-(h/2) < X < x + (h/2)]$$

$$= \lim_{h \to o} h^{-1} E[I(-\frac{1}{2} < \frac{X-x}{h} < \frac{1}{2})]$$

where $F(x) = P[X < x]$ is the cumulative probability distribution function and $I(a < X < b)$ is 1 or 0 according to whether X is in the interval (a,b) or outside this interval.

Let $x_1,\ldots,x_n$ be independent identically distributed (i.i.d.) observations on X. Then an obvious consistent estimator of f(x) is given by

$$f_n = f_n(x) = \frac{F_n(x+h/2) - F_n(x-h/2)}{h} \tag{2.2}$$

$$= h^{-1} n^{-1} \sum_{t=1}^{n} I(-\frac{1}{2} < \frac{x_t - x}{h} < \frac{1}{2})$$

where $F_n(x)$ is the empirical distribution of $x_1,\ldots,x_n$ given by

$$F_n(x) = \frac{\text{number of } x_1,\ldots,x_n \text{ that are} \le x}{n} \qquad (2.3)$$

and the window-width $h = h_n$ is a positive function of the sample size n which goes to zero as $n \to \infty$.

To see that (2.2) is an obvious consistent estimator of f, notice that for each fixed h, $h^{-1} n^{-1} \sum_{t=1}^{n} I(-\frac{1}{2} < \frac{x_t - x}{h} < \frac{1}{2})$ is a consistent (as $n\to\infty$) estimator of $h^{-1} E\{I(-\frac{1}{2} < \frac{x_t - x}{h} < \frac{1}{2})\} = h^{-1}\{F(x + \frac{h}{2}) - F(x - \frac{h}{2})\}$, which in turn approaches to the density f(x) as $h\to 0$ by the definition of a density function. Rosenblatt (1956); taking $h = h_n$, a positive function of n approaching to zero as $n\to\infty$, proved this and various other properties of f_n. He further noted that if the non-negative function $I(w) = I(-\frac{1}{2} < w < \frac{1}{2})$ in (2.2) satisfying $\int I(w)dw = 1$ is replaced by any other non-negative function K(w) satisfying $\int K(w)dw = 1$ the consistency of property of f_n remains invariant. For this reason he introduced the more general "kernel" estimator

$$\hat{f}(x) = h^{-1} n^{-1} \sum_{t=1}^{n} K(\frac{x_t - x}{h}) , \qquad (2.4)$$

which is asymptotically unbiased and consistent. (The choice of h is discussed in Section 2.5.) It appears that Whittle (1958), independently of Rosenblatt, has derived a somewhat similar method of density estimation.

Among the various methods of univariate density estimation are: the polynomial series (e.g., Schwartz (1967), Kronmal and Tarter (1968) and Watson (1969)), the maximum likelihood (e.g., Wegman (1970)), the histogram (Van Ryzin (1973)) and the nearest neighbourhood methods (e.g., Moore and Yackel (1977)). The most widely used one is the kernel method of estimation, and this has been considered throughout the paper.

Parzen (1962) extended Rosenblatt's estimator to cases where the weight function need not be non-negative. For any Borel measurable function K(w) satisfying $\int K(w)dw = 1$, $\sup_{-\infty<w<\infty} |K(w)| < \infty$, $\int |K(w)|dw < \infty$ and $\lim_{|w|\to\infty} |wK(w)| = 0$, he showed that $\hat{f}(x)$ in (2.4) is asymptotically unbiased and mean square consistent at every continuity point of f. The asymptotic distribution of $(\hat{f}(x) - E\hat{f}(x))/\sqrt{V}(\hat{f}(x))$ is shown to be standard normal.

Some further asymptotic properties of $\hat{f}(x)$ have been investigated in Nadarya (1964) and Schuster (1976).

A paper by Johns and Van Ryzin (1972) showed how to construct estimators whose mean squared errors (MSE) converge to zero, as $n\to\infty$, more quickly, as suggested by Parzen (1962) and Bartlett (1963). A method, different from that of Johns and Van Ryzin, was proposed by Schuncany and Sommers (1976). Their restimators, which turn out to be linear combinations of the original estimators (Rosenblatt (1956); Parzen (1962)), utilize the generalized jackknife method of Schucany, Gray and Owen (1971).

Motivated by Rosenblatt (1956) and Parzen (1962), the following estimator of f

$$f_n^*(x) = n^{-1} \sum_{t=1}^{n} h_t^{-1} K\left(\frac{x_t - x}{h_t}\right) , \qquad (2.5)$$

closely related to (2.4), was introduced by Wolverton and Wagner (1969) and apparently independently by Yamato (1971). Notice that, if additional data become available, then to update the estimator (2.4), incorporating all the (n+1) observations, one has to recompute all the n+1 functions $K((x_t - x)/h_{n+1}), t=1,\ldots,n+1$, a rather tedious job. In contrast, the estimator in (2.5) is recursive in the sense that it can be updated by the formula

$$f_{n+1}^*(x) = (n+1)^{-1} [n\, f_n^*(x) + h_{n+1}^{-1} K\left(\frac{x_t - x}{h_{n+1}}\right)] . \qquad (2.6)$$

The estimator f_n^* is also asymptotically unbiased and mean squared consistent. Various other asymptotic properties of $f_n^*(x)$ have been discussed in Davies (1973), Deheuvels (1974b), Davies and Wegman (1975) and Carroll (1976).

The kernel estimator in (2.4) has been extended to the case of dependent time series observations case by Robinson (1983) and Singh and Ullah (1985), among others.

Bhattacharya (1967) and later Schuster (1969) used Rosenblatt's estimates of f to construct estimates of the r^{th} order, $r \geq 0$, derivatives $f^{(r)}$ of f. This is given by

$$\hat{f}^{(r)}(x) = \frac{1}{nh^{r+1}} \sum_{t=1}^{n} K^{(r)}\left(\frac{x_t - x}{h}\right) \qquad (2.7)$$

where $K^{(r)}$ is the r^{th} order derivative of K. Notice that (2.7) is the r^{th} order derivative of (2.4). Alternative estimators and their properties have been developed in Singh (1974, 1977, 1979b, 1981). Singh (1974, 1977, 1980a, 1980b) also developed estimators of the average of the densities or their derivatives.

2.2 Multivariate Case

The kernel method of estimation of a univariate density function, introduced by Rosenblatt (1956) and studied in detail by Parzen (1962) and Bartlett (1963), was first generalized to the multivariate density function by Cacoullos (1966). Let x_t, $t=1,\ldots,n$ be n independent m X 1 random vectors generated from an unknown m-variate density function. Consider $\mathcal{K}$ to be a class of all Borel-measurable real valued bounded functions K on the m-dimensional Euclidean space R^m such that

$$\int K(w)dw = 1, \qquad \int |K(w)|dw < \infty , \tag{2.8}$$

$$\|w\|^m |K(w)| \to 0 \quad \text{as } \|w\| \to \infty ,$$

where $\|w\|$ is the usual Euclidean norm of w in R^m. Then Cacoullos estimated f at a point x in R^m by

$$\hat{f}(x) = n^{-1} h^{-m} \sum_{t=1}^{n} K\left(\frac{x - x_t}{h}\right) \tag{2.9}$$

where as before $h = h_n \to 0$ as $n\to\infty$. He showed that the estimator $\hat{f}(x)$ is asymptotically unbiased and mean square consistent. Further for $x_1,\ldots,x_j$, distinct continuity points of f in R^m, the vector $(\hat{f}(x_1),\ldots,\hat{f}(x_j))$ is asymptotically j-variate normal. Cacoullos work was extended to estimation of partial (or any mixed partial) derivatives of a density by Singh (1978, 1981). In Singh (1981), the speeds of various convergences of the estimators to the true value are also examined.

In the multivariate case, the recursive estimator corresponding to $f_n^*(x)$ in (2.5) has been proposed in Singh and Ullah (1984). In their paper, Singh and Ullah consider the case when $h = h_{it}$, $i=1,\ldots,m$ and $t=1,\ldots,n$. They show, among other things, that the recursive estimator is more efficient compared to $\hat{f}(x)$ in (2.9). For the case of dependent observations (especially in the time series context) the estimators of type (2.9) and their properties have been analyzed in Robinson (1983) and Singh and Ullah (1985). For some other work on the estimation of a multivariate density, see Loftsgaarden and Quesenbury (1965), and Epanechniko (1969), among others.

2.3 Estimators of Marginal and Conditional Densities

Let us write the m components of the m X 1 vector x_t of section (2.2) as

$$x_t = [y_t', z_t']', \qquad t=1,\dots,n \qquad (2.10)$$

where y_t is a p X 1 vector and z_t is a q X 1 vector such that p + q = m. Similarly, as in Section 2.2., let K_1 and K_2 be p and q variate functions obtained after integrating $K(w_1,\dots,w_m)$ with respect to $(w_{p+1},\dots,w_m)$ and with respect to $(w_1,\dots,w_p)$ respectively. Further, consider $x = [y'z']'$ be a point in R^m at which the density is to be estimated.

Now the estimator of the joint density f at x from (2.9) is

$$\hat{f}(x) = n^{-1} \sum_{t=1}^{n} h^{-m} K\left(\frac{x_t - x}{h}\right) = \hat{f}(y,z) \,. \qquad (2.11)$$

Using this we can write the marginal density of z_t at z as

$$\hat{f}(z) = \int \hat{f}(y,z)dy = n^{-1} \sum_{t=1}^{n} h^{-m} \int K\left(\frac{x_t - x}{h}\right)dy \qquad (2.12)$$

$$= n^{-1} \sum_{t=1}^{n} h^{-q} K_2\left(\frac{z_t - z}{h}\right)$$

The marginal density of y can be similarly written.

Next, the estimator of the conditional density of y_t given $z_t = z$ can be obtained as

$$\hat{f}(y|_z) = \frac{\hat{f}(y,z)}{\hat{f}(z)} = \frac{\sum_{t=1}^{n} h^{-m} K\left(\frac{y_t - y}{h}, \frac{z_t - z}{h}\right)}{\sum_{t=1}^{n} h^{-q} K_2\left(\frac{z_t - z}{h}\right)} \qquad (2.13)$$

where $\hat{f}(y,z)$ and $\hat{f}(z)$ are as given in (2.11) and (2.12) respectively.

In an interesting special case the joint kernel function is

$$K(w_1,\dots,w_m) = k_1(w_1)\dots k_p(w_p)k_{p+1}(w_{p+1})\dots k_m(w_m) \qquad (2.14)$$

$$= \prod_{i=1}^{p} k_i(w_i) \prod_{j=1}^{q} k_{p+j}(w_{p+j}),$$

where k_j are univariate symmetric (around zero) Borel-measurable functions. Notice that k_j, for j=1,...,m, can be identical. Taking $k_1 = \dots = k_m = k$ (say) we can write (2.14) as

$$\hat{f}(y|_z) = \frac{\sum_{t=1}^{n} h^{-m} \{ \prod_{i=1}^{p} k(\frac{y_{ti} - y_i}{h}) \prod_{j=1}^{q} k(\frac{z_{tj} - z_j}{h}) \}}{\sum_{t=1}^{n} h^{-q} \prod_{j=1}^{q} k(\frac{z_{tj} - z_j}{h})} \tag{2.15}$$

where y_{ti}, $i=1,\ldots,p$ and p elements of y_t and z_{tj}, $j=1,\ldots,q$, are q components of z_t.

These estimators of conditional and marginal densities are asymptotically unbiased and mean squared consistent. These, and other distributional properties, have been discussed in Singh and Ullah (1984) for independent observations, and in Singh and Ullah (1985) for dependent observations.

2.4 Properties of Estimators and Confidence Intervals

Here we state the main results on unbiasedness, variance, consistency and distribution of the estimators $\hat{f}$ of f and $\hat{f}(y|_z)$ of $f(y|_z)$. The proofs of the results are given in Singh and Ullah (1984) for independent observations, and in Singh and Ullah (1985) for dependent observations.

Asymptotic Unbiasedness: Let x be a point in R^m at which f is continuous. Then for $K \in \mathcal{K}$

$$B(x) = (E\hat{f}(x) - f(x)) = o(1) \tag{2.16}$$

and

$$\text{Sup}_x |B(x)| = o(1) \; . \tag{2.17}$$

Variance: At every continuity point x of f,

$$V(\hat{f}(x)) = (nh^m)^{-1} f(x) \int K^2 \{1 + o(1)\} \; . \tag{2.18}$$

Consistency: If $(nh^m)^{-1} = o(1)$, then at every continuity point x of f

$$\hat{f}(x) - f(x) = o_p(1) \; . \tag{2.19}$$

For any integer $r \geq 0$, let $\mathcal{K}_r$ be the class of real valued Borel-measurable bounded functions K on R^m such that

$$\begin{aligned} \int u_1^{i_1} \ldots u_m^{i_m} K(u_1,\ldots,u_m) &= 1 \text{ if } i_1 + \ldots + i_m = 0 \\ &= 0 \text{ if } 0 < i_1 + \ldots + i_m < r \end{aligned} \tag{2.20}$$

$\int\|u\|^i|K(u)| < \infty$ for $i = 0$ and r

$\|u\|^m|K(u)| \to 0$ as $\|u\| \to \infty$.

For example, for $r = 0,1,2$ m variate standard normal density belongs to $\mathcal{K}_r$, and so is the function $K(u_1,\ldots,u_m) = 2^{-m}\Pi_{j=1}^m I(-1 < u_j < 1)$, where $I(\cdot)$ stands for the indicator function.

If for some $r \geq 1$, r^{th} order partial derivatives of f are continuous at x, then for K in $\mathcal{K}_r$ (see (2.20))

$$B(x) = O(h^r) \text{ and } \sup_{x\epsilon R^m}|B_n(x)| = O(h^r) . \tag{2.21}$$

Further, if (2.18) and (2.21) hold then

$$\hat{f}(x)f-(x) = O_p(h^r+(nh^m)^{-1/2})$$

$$\sup_x|\hat{f}(x)-f(x)| = O_p(h^r+(nh^m)^{-1/2}) . \tag{2.22}$$

Note also that when h is chosen so that

$$h \sim n^{-1/(2r+m)} \tag{2.23}$$

then from (2.22)

$$\hat{f}(x)-f(x) = O_p(n^{-r/2r+m}) \text{ and } \sup_x|\hat{f}(x)-f(x)| = O_p(n^{-r/2r+m}) . \tag{2.24}$$

It has been shown in Singh and Ullah (1985) that if the characteristic function of the function K involved in the definition of $\hat{f}$ is absolutely integrable then $\hat{f}(x)$ is uniformly weak consistent, that is,

$$P[\sup_x|\hat{f}(x)-f(x)| > \epsilon] = o(1) \text{ for every } \epsilon > 0 . \tag{2.25}$$

Asymptotic Normality: If for some $r \geq 1$, r^{th} order partial derivatives of f are continuous at x, and with this r, K is taken from $\mathcal{K}_r$, then choosing h as in (2.23) we have

$$(nh^m)^{1/2}(\hat{f}(x)-f(x)) \sim N(0,f(x)\int K^2) . \tag{2.26}$$

From (2.26) 100(1-β)% C.I. for f(x) is

$$\hat{f}(x) \pm z_{\beta/2}(nh^m)^{-1/2}[f(x)\int K^2]^{1/2} \tag{2.27}$$

where $z_{\beta/2}$ is such that if Z is the univeriate standard normal

$$P[-z_{\beta/2} < Z < z_{\beta/2}] = 1-\beta . \tag{2.28}$$

Asymptotic Normality of $\hat{f}(y|_z)$: If for some $r \geq 1$, r^{th} order partial derivatives of f are continuous at x, and h is as in (2.23) we have

$$(nh^m)^{1/2}[\hat{f}(y|_z)-f(y|_z)] \sim N(0, \frac{f(y,z)}{[f(z)]^2}\int K^2) \qquad (2.29)$$

where $\hat{f}(y|_z)$ is as given in (2.13).

The 100(1-β)% C.I. for $f(y|_z)$ can easily be written by using (2.29).

We note that for dependent observations strict stationarity of the process x is assumed. Further for proving asymptotic properties certain conditions on strongly mixing coefficients are required. For details see Singh and Ullah (1985).

2.5 On the Choice of the Kernel K and the Window-Width Function h

First for the univariate case we note from the work of Singh (1979b, 1981) and (2.21) that

$$B(x) = h^r f^{(r)} k_r = O(h^r), \qquad r \geq 1 \qquad (2.30)$$

where $f^{(r)}$ represents r^{th} order partial derivatives of f and $k_r = 1/r! \int u^r K(u)du$; K belonging to $\mathcal{X}_r$ in (2.20) such that the first r-1 moments of K are zero.

Using (2.18) and (2.30) we can write the mean squared error (MSE) of $\hat{f}(x)$ as

$$MSE(\hat{f}(x)) \sim [h^{2r}(f^{(r)}k_r)^2 + n^{-1}h^{-1}f\int K^2] \qquad (2.31)$$

An examination of (2.31) indicates that, for given h, we should choose K so that k_r^2 and $\int K^2$ are not too large (in fact smaller the better). Such a choice of K, however, should satisfy the following moment conditions (see (2.20))

$$\int u^j K(u) = 1 \qquad \text{if } j = 0 \qquad (2.32)$$
$$= 0 \qquad \text{if } j = 1,\ldots,r-1 .$$

To look into K's which satisfy (2.32) consider

$$K(u) = (\sum_{i=o}^{r-1} a_i u^i)\phi(u), \qquad \phi(u) \sim N(0,1) . \qquad (2.33)$$

Then it can be verified that

$$K(u) = \phi(u) \qquad \text{for } r = 1 \text{ or } 2 \qquad (2.34)$$

$$= \frac{1}{2}(3-u^2)\phi(u) \qquad \text{for } r = 3 \text{ and } 4$$

$$= [2\phi(u)-2^{-1}\pi^{-1/2}e^{-u^2/4}] \qquad \text{for } r = 3 \text{ and } 4 .$$

However, if $\phi(u) = \frac{1}{2} = \frac{1}{2} I(-1 < w < 1)$ then

$$K(u) = \frac{1}{2} \qquad \text{for } r = 0,1,2 \qquad (2.35)$$

$$K(u) = \frac{(9-15u^2)}{4}\frac{1}{2} .$$

For a given kernel, we now consider the choice of window-width h. An examination of MSE in (2.31) indicates that as h decreases the $(\text{bias})^2$ decreases but the variance increases. Therefore, we should try to choose h in such a way that the bias and variance both remain under control. Such a choice of h may be taken as $cn^{-1/(1+2r)}$ since it makes both the $(\text{bias})^2$ and variance, hence MSE, to be of order $O(n^{-2r/(1+2r)})$ provided c does not depend on n. Using $h = cn^{-1/(1+r)}$ we note from (2.31) that the optimal choice of c is

$$c_o = \left[\frac{1}{2r}\left(\frac{f}{(f^{(r)})^2}\right)\frac{\int K^2}{k_r^2}\right]^{1/(1+2r)} . \qquad (2.36)$$

However, this requires some knowledge of the ratio $f/(f^{(r)})^2$ and thus it is not useful in practice.

Another point to note is that the $V(\hat{f}(x))$ will become inflated whenever $V[K(\frac{x_t - x}{h})]$ is large which occurs when $V(x_t) = \sigma^2$ (say) is large. This effect of σ^2 on $V[K(\frac{x_t - x}{h})]$ will be eliminated if h is proportional to σ. Therefore, a good choice of h is

$$h\alpha\sigma n^{-1/(1+2r)} = c\sigma n^{-1/(1+2r)} \qquad (2.37)$$

We recommend (2.37) especially when σ is large. A too small σ will again inflate the variance of $\hat{f}$, as it can be ssen from (2.18).

Next, we observe that the speed noted above for the MSE approaches to $O(n^{-1})$ as r gets large. However, for large r, construction of a kernel satisfying the moment conditions set out earlier will not be an easy job, and the estimator based on such a kernel will be difficult to compute.

If x is m X 1 vector then the h can be chosen as

$$h_i \alpha \sigma_{x_i} n^{-1/(m+2r)} \quad , \; i=1,\ldots,m \; , \qquad (2.38)$$

and

$$K(u_1,\ldots,u_m) = \prod_1^m K(u_i) \; , \qquad (2.39)$$

where $K(u_i)$ is as given in the univariate case discussed before.

3. APPLICATIONS

In this section we investigate the applications of density estimates, given in Section 2, for the various problems in econometrics.

3.1 Estimation of Conditional Mean (Regression Function)

Let y and $z_1,\ldots,z_q$ be a set of q + 1 random variables. The conditional expectation of y given the values of $z_1,\ldots,z_q$ is then given by

$$E(y|_{z_1,\ldots,z_q}) = M(z_1,\ldots,z_q) = M(z) \qquad (3.1)$$

The function M( ) shows how the average values of y change with a change in the value of $z_1,\ldots,z_q$. This function plays a significant role in econometrics for the purposes of prediction and testing economic theories. However, we note that g is known only if either the data generating process (joint density $f(y,z_1,\ldots,z_q)$) is known or the true $f(y|_{z_1,\ldots,z_q})$ is known. Since these are rarely, if ever, known the econometricians have invariably specified, a priori,

$$E(y|_{z_1,\ldots,z_q}) = M(z_1,\ldots,z_q) = z_1\beta_1 + \ldots z_q\beta_q \qquad (3.2)$$

and labelled it "the linear regression function" or "the linear conditional model". The least square theory is then used for estimation and prediction purposes.

A useful alternative proposed here is to estimate directly the multivariate density $f(y,z_1,\ldots,z_q)$ and the marginal density $f(z_1,\ldots,z_q)$ by the methods discussed in Sections 2.2 and 2.3, and then estimate (3.2). For example, using (2.13) we can get

$$\hat{f}(y|_z) = \frac{\hat{f}(y,z)}{\hat{f}(z)} = \frac{h^{-1}\sum_{t=1}^{n} K(\frac{y_t - y}{h}, \frac{z_t - z}{h})}{\sum_{t=1}^{n} K_2(\frac{z_t - z}{h})} \tag{3.3}$$

and

$$\hat{E}(y|_{z_1,\ldots,z_q}) = \int y\hat{f}(y|_{z_1,\ldots,z_q})dy \tag{3.4}$$

$$= \frac{\sum_{t=1}^{n} y_t K_2(\frac{z_t - z}{h})}{\sum_{t=1}^{n} K_2(\frac{z_t - z}{h})} = \sum_{t=1}^{n} y_t r_t(z) = \hat{M}(z)$$

where

$$r_t(z) = K_2(\frac{z_t - z}{h}) / \sum_{t=1}^{n} K_2(\frac{z_t - z}{h}) . \tag{3.5}$$

The $\hat{E}(y|_{z_1,\ldots,z_q})$ in (3.4) can be used for econometric analysis. For example, the forecast of y for a given value of z can be obtained from (3.4). Also the change of y due to a unit change in, say z_1, can be determined by calculating partial derivatives of (3.4). See the illustrative examples in Section 4 for the estimates of (3.4).

We note that the nonparametric estimator (conditional model) in (3.4), unlike (3.2) is obtained without making assumptions about the functional form, the joint density of y and z's, and non-stochastic behaviour of z's. Further, since (3.4) has been obtained by estimating joint density we do not require to check up the weak exogeneity of z. Thus the four problems of parametric inference indicated in the beginning of Section 1 are absent in the nonparametric estimator (3.4).

The following result has been proved in Singh and Ullah (1985).

<u>Consistency and Asymptotic Normality:</u> <u>Let the r^{th} order partial derivatives of</u> $\int y^i f(y,z)dy$ (i=0,1,2) <u>be continous at</u> z. <u>Then taking</u>

$$h \alpha n^{-1/(2r+m-1)} \tag{3.6}$$

<u>we have</u>

$$\hat{M}(z) = M(z) + O_p(n^{-\frac{r}{2r+m-1}}) \qquad (3.7)$$

and

$$(nh^{m-1})^{1/2} (\hat{M}(z)-M(z)) \sim N(0, \frac{V(y|_z)\int K^2}{f(z)}) , \qquad (3.8)$$

where $V(y|_z)$ is the conditional variance of y given z.

From (3.8), 100(1-β)% C.I. for M(z) is

$$\hat{M}(z) \pm z_{\beta/2}(nh^{m-1})^{-1/2} [\frac{V(y|_z)\int K^2}{f(z)}]^{1/2} . \qquad (3.9)$$

In practice, (3.9) can be used by replacing f(z) by its consistent estimator $\hat{f}(z)$ given in (2.12), and $V(y|_z)$ by its consistent estimator given in Section 3.2.

a. Estimation of Linear Probability Model

The estimator (3.4) can also be used in the context of a regression model in which the dependent variable y is a binary variable taking the value $y_t = 1$ with probability p_t if the event occurs and $y_t = 0$ with probability $1 - p_t$ otherwise, t=1,...,n. Examples of this are participation in the labor force, decision to marry, bankruptcy, etc. Note that for the binary variable y_t

$$Ey_t = p_t \quad \text{and} \quad V(y_t) = p_t(1-p_t) . \qquad (3.10)$$

In empirical econometrics work various assumptions regarding p_t have been made. Some of these are as given below

(i) Probit: $p_t = (2\Pi)^{-1/2} \int_{-\infty}^{z_t\beta} e^{-w^2/2} dw$ and

(ii) Logit: $p_t = (1 + e^{-z_t\beta})^{-1}$

so that p_t in these cases become cumulative probability distribution function $F(z_t\beta) = E(y_t|_{z_t})$, z_t is a q X 1 given vector and β is a q X 1 vector of parameters. The likelihood function $L = \prod_{t=1}^{n} p_t^{y_t} (1-p_t)^{1-y_t}$ is then written and the parameters are estimated.

It is clear from (3.10) that the specification of the linear probability model amounts to specifying the probability $p_t = E(y_t|_{z_t})$ by a suitable cumulative probability density. An alternative is to consider the nonparametric approach discussed in Section 2. That is, if our interest is to estimate the conditional expectation with respect to vector z_t, $\hat{p}_t = \hat{E}(y_t|_{z_t})$, then we can verify that it is as given in (3.4). Note, however, that in our present model y_t is either 1 or 0. Thus, $\hat{p}_t$ from (3.4) will be between 0 and 1.

The qualitative response models discussed in Mcfadden (1984) can also be similarly analyzed.

b. Censored and Truncated Models

The statistical literature on the estimation of censored normal and truncated normal distribution is very long (see Cohen (1950), Hald (1949) and Halperin (1952)). In econometrics, censored normal models have been used extensively by Tobin (1958), Amemiya (1973), Heckman (1976) and Fair (1977) and truncated normal models have been used by Hausman and Wise (1976, 1977) among others.

Suppose $f(y^*,z)$ is the joint density function of y^* and z. Let $y_1^*,\ldots,y_n^*$ and $z_1,\ldots,z_n$ be the samples of size n. For y^* we record only those values which are greater than a constant c. For those values of $y^* \leq c$, we record the value c. Thus, for $t=1,\ldots,n$,

$$y_t = y_t^* \quad \text{if } y_t^* > c \qquad (3.11)$$
$$\quad = c \quad \text{otherwise.}$$

The resulting sample $y_1,\ldots,y_n$ is said to be a censored sample.

For the above case

$$E(y|_z) = cP[y=c|_z] + \int_c^\infty yf(y|_z)dy \qquad (3.12)$$

or

$$E(y|_z) = c\,\frac{f[y=c,z]}{f(z)} + \int_c^\infty y\,\frac{f(y,z)}{f(z)}\,dy\,. \qquad (3.13)$$

The model in (3.13) can then be analyzed by using the nonparametric estimates of the marginal and joint densities given in Section 2.3.

Now consider the truncated model. Suppose $y_1,\ldots,y_n$ is a sample drawn from the truncated population of $y < c_o$. Then the truncated model is

$$f(y_t) = f(y_t | y_t < c_o) = \frac{f(y_t)}{F(c_o)} \tag{3.14}$$

where $F(c_o)$ is the cumulative distribution of $f(y)$. And in the case of two variables y_t and z_t

$$f(y_t |_{z_t}) = \frac{f(y_t, z_t |_{y_t} \le c_o)}{f(z_t)} = \frac{f(y_t, z_t)}{F(c_o) f(z_t)} \tag{3.15}$$

Again for prediction ($E(y|_z)$) and other econometric analyses the estimates of the marginal and conditional densities given in Section 2.3 can be used. This would overcome the specification of $E(y|_z) = z\beta$ as well as the normality assumption used by Hausman and Wise (1976, 1978) or the assumption of Edgeworth density by Lee (1982). Note that Cosslett (1978, 1980) suggests a pseudo-nonparametric procedure for the above model.

3.2 Estimation of Conditional Variance (Conditional Heteroscedasticity)

Let us write the conditional variance of y given $z_1,\ldots,z_q$ as

$$V(y|_{z_1,\ldots,z_q}) = V(z) = E(y^2|_z) - [E(y|_z)]^2 . \tag{3.16}$$

Then a consistent nonparametric estimator (see Singh and Ullah (1985)) is given by

$$\hat{V}(z) = \Sigma y_t^2 r_t(z) - \hat{M}^2(z) \tag{3.17}$$

where $r_t(z)$ and $\hat{M}(z) = \Sigma y_t r_t(z)$ are as given in (3.5) and (3.4), respectively. The higher order conditional moments of y can similarly be estimated.

The estimation of variability is of interest due to various reasons. First, there are many economic models in which the variability (risk term) appears as a regressor and, second, variability in economic variables such as inflation and interest rates is in itself of interest to the policymakers. For details, see Friedman (1977). In addition there are many economic models in which the conditional variance of the dependent variable is heteroscedastic. Thus a nonparametric estimator (3.17), which does not use any functional form of heteroscedasticity, is useful for the efficient generalized least squares estimation of the model.

3.3 Model Adequacy and Other Tests

It was noted in Section 3.1 that the nonparametric estimator $\hat{E}(y|_z)$ in (3.4) is obtained without making assumptions about the weak exogeneity of z, functional form, and the joint density of y and z. Thus the nonparametric residuals

$$y_t - \hat{E}(y_t|_{z_t}) = \hat{u}_t, \qquad t=1,\ldots,n$$

are robust and they can be used to perform meaningful diagnostic tests for the adequacy of the model

$$y_t = E(y_t|_{z_t}) + u_t .$$

This can be done by simply using nonparametric residuals $\hat{u}_t$ and the fitted values $\hat{E}(y_t|_{z_t})$, instead of non-robust least squares residuals and fitted values, in various diagnostic tests for normality, heteroscedasticity, serial correlation, exogeneity, misspecification and encompassing (non-nested) given in Pagan (1983) and Ullah (1985). A point to be noted here is that the nonparametric residual $\hat{u}_t$ is such that

$$\hat{u}_t - u_t = -(E(y_t|_{z_t}) - \hat{E}(y_t|_{z_t})) = O_p(n^{-\frac{r}{2r+m-1}})$$

because of (3.7).

Since most of the diagnostic tests may be non-robust under misspecifications a better alternative will be to use the results of Section 2 directly. For example, the normality can be checked up by estimating $f(\hat{u})$ and calculating its C.I. from (2.27). Similarly, the misspecification, heteroscedasticity, and serial correlation can be analyzed by using nonparametric estimates of $E(\hat{u}|z)$, $V(\hat{u}|z)$ and $\mathrm{cov}(\hat{u}_t,\hat{u}_{t-1}|z)$, respectively. Some other tests can also be performed and these are given below.

a. Test of Independence

For the test of independence (see Hausman (1978)) we can estimate the conditional and marginal densities, $f(z|\hat{u})$ and $f(z)$, respectively by using the methods of Section 2.3 and calculate

$$d^2 = \sum_{u}^{\ell} \sum_{z}^{m} \hat{d}^2(z|u); \qquad \hat{d}(z|u) = (\hat{f}(z|u))^{1/2} - (\hat{f}(z))^{1/2} . \qquad (3.18)$$

The statistic d^2 is then the statistic for checking the independence of $\hat{x}$ and u. Notice that d^2 is Bhattacharya's (1967) distance measure which satisfies all the properties of a metric. An alternative to Bhattacharya's distance would be to use Kulback-Leibler information divergence measure in which $\hat{d}(z|u) = -(\log \frac{\hat{f}(z)}{\hat{f}(z|u)})\, \hat{f}(z|u)$. For details on divergence measures see Burbea and Rao (1982) and Ullah (1983).

It is our conjecture that d^2 may follow chi-square asymptotically. The work on this will follow in a future paper.

b. Testing Causality

In most economic models not only is a cause and effect relationship assumed to hold but, furthermore, the direction of causality is also taken to be known. The truth is, however, that in non-experimental subjects like economics it is difficult to find convincing evidences in favour of such assumptions. In view of this, following the work of Wiener, Granger (1969) first formalized the idea of causality. The essence of Granger's causality is that z does not cause y if

$$f(y_t|_{\Omega_{t-1}}) = f(y_t|_{\Omega^y_{t-1}}) \qquad (3.19)$$

where $\{\Omega_{t-1} = \Omega^y_{t-1}, \Omega^z_{t-1}\}$ is the information set consisting of past values of y as well as z. Note that (3.19) is the "causality with respect to the particular Ω_{t-1} used." Since the conditional distribution functions in (3.19) are unknown, a testable definition has been used in terms of a summary statistic, viz, linear predictions. More precisely z causes y if

$$MSE(y|_{\Omega_{t-1}}) < MSE(y|_{\Omega^y_{t-1}}) . \qquad (3.20)$$

An alternative but equivalent test proposed in the literature is to regress the current value of y on the lagged values of z as well as y and testing for the coefficients of lagged values of z are zero by the F-test. If the hypothesis is accepted z is said to be not causing y in Granger's sense.

Note that the simplified Granger's causality test would be useful, that is the F-test proposed would have power, if there are no misspecifications in the specified variables z and y and there is no misspecification in the error term of the regression indicated above. Further, although the definition in (3.20) is simple, it is a long way from the rather general definition started with in (3.19). The true causality may be missed, or spurious causality observed, because of these simplifications. Thus, again an alternative is to use nonparametric methods to estimate the conditional density $f(y_t | z_1, \ldots, z_{t-1}, y_1, \ldots, y_{t-1})$ and the marginal density $f(y_t | y_1, \ldots, y_{t-1})$ and compare them. The variable z does not cause y if these two estimated densities are the same, that is

$$\hat{f}(y_t | z_1, \ldots, z_{t-1}, y_1, \ldots, y_{t-1}) = \hat{f}(y_t | y_1, \ldots, y_{t-1}) \qquad (3.21)$$

We can also check whether or not there is a significant difference between the conditional and marginal densities by using the statistic

$$d^2 = \sum_z^{\ell} \sum_y^{m} d^2(y_t | z_{t-1}, y_{t-1}) \; ; \qquad (3.22)$$

$$d(y_t | z_{t-1}, y_{t-1}) = \hat{f}(y_t | z_1, \ldots, z_{t-1}, y_1, \ldots, y_{t-1}) - \hat{f}(y_t | y_1, \ldots, y_{t-1}).$$

We can also check for the instantaneous independence of z and y by using

$$d^2 = \sum_z^{\ell} \sum_y^{m} d^2(y_t | z_t); \qquad d(y_t | z_t) = \hat{f}(y_t | z_t) - \hat{f}(y_t) \qquad (3.23)$$

c. <u>Non-Nested Model Selection</u>

Let x, y and z be three economic variables such that there are two data generating processes for y, that is $f(y|_x)$ and $f(y|_z)$. If the maintained hypothesis is $f(y|_z)$ then the non-nested model selection problem is to see if $f(y|_z)$ is significantly different from $f(y|_x)$. Usually the parametric specification in terms of regression models is used in order to test such an hypothesis (see, e.g., Davidson and MacKinnon (1981)). We propose the estimation of $f(y|_x)$ and $f(y|_z)$ first and then the use of

$$d^2 = \sum_y^m \{g_1^{1/2}(y) - g_2^{1/2}(y)\}^2; \quad g_1(y) = \sum_x^{\ell} \hat{f}(y|_x), \; g_2(y) = \sum_z^{\ell} \hat{f}(y|_z)$$

where ℓ is ℓ-number of points of x and z. Again, as in the case a, one can use Kulback-Leibler divergence measure $d^2 = \sum_y^m \log \frac{\hat{g}_2(y)}{\hat{g}_1(y)} \hat{g}_1(y)$.

3.4 Finite Sample Econometrics

There is a large literature in econometrics on deriving the exact and approximate densities and moments of various econometric estimators (see, e.g., Phillips (1983), among others). These works are of great importance for proper inference in finite sample situations, particularly since the concept of a "large" sample is fuzzy for practical situations. However, despite its great importance, work in this area has so far not been very useful because of the complicated expressions of exact results and their nonlinear dependence on unknown parameters.

We propose a nonparametric Monte-Carlo integrated approach here. Suppose, based on the 5000 random samples of size 10 we generate 5000 observations on the estimator of a parameter ($\hat{\beta}$) in a parametric specification. Using these observations on $\hat{\beta}$ we can then easily estimate the unknown exact density of $\hat{\beta}, f(\hat{\beta})$, by the nonparametric method discussed in 2.1, say $\hat{f}(\hat{\beta})$. Moments of the estimators, based on $\hat{f}(\hat{\beta})$ can then be analyzed easily. The results can be compared with the Monte-Carlo results as well as the exact results, available in the literature. Notice that the procedure can be extended to the estimation of multivariate or marginal density of a parameter vector β appearing in an econometric model. Similarly, kernel density estimation can be integrated with the bootstrapping approach of Efron (1979).

4 ILLUSTRATIVE EXAMPLES

Here we present some examples, based on Monte-Carlo as well as real world data sets, to illustrate the techniques of Section 2 in the context of econometric problems discussed in Sections 3.1 and 3.2. The illustrations for the issues in Sections 3.3 and 3.4 are beyond the scope of this paper, and they will be subjects of a future study.

4.1 Estimation of Conditional Mean and Variance (Monte-Carlo)

The objective of the experiment here is to verify how the nonparametric estimates of conditional mean (regression function) and variance (heteroscedasticity) perform when the true forms of the conditional mean and variance are known, and the data are generated from a known population.

Consider the true model as a quadratic regression function given by

$$y_t = \beta_o + \beta_1 z_t + \beta_2 z_t^2 + u_t, \qquad t=1,\ldots,n \tag{4.1}$$

where y is a dependent variable, z is an exogenous variable, β's are regression parameters, and u is the disturbance term such that

$$u_t \sim N(0,\sigma^2) \tag{4.2}$$

From (4.1) and (4.2), the parametric forms of the conditional mean and variance are

$$E(y_t|_{z_t}) = \beta_o + \beta_1 z_t + \beta_2 z_t^2 \quad \text{because } E(u_t|_{z_t}) = Eu_t = 0 \tag{4.3}$$

$$V(y_t|_{z_t}) = V(u_t|_{z_t}) = V(u_t) = \sigma^2 .$$

To generate the observations on y we (i) specify $\beta_o = 5$, $\beta_1 = 1$, $\beta_2 = 3$, (ii) choose n = 50 values of z of an economic variable and (iii) generate random samples of sizes n = 50, 100, 400 from $u \sim N(0,1)$. This gives samples of sizes n = 50, 100, 400 for the y from (4.1). Note that the values of z for n = 100 and 400 are generated by repeating its 50 values.

The nonparametric estimates of $E(y|_z)$, given in (3.4), are plotted in Figures 1 and 2, respectively for n = 100 and 400. The kernel chosen for this purpose was

$$K\left(\frac{z_t - z}{h}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z_t - z}{h}\right)^2} \tag{4.4}$$

where $h = s_z n^{-1/5}$, $s_z^2 = \sum_1^n (z_t - \bar{z})^2/n$. It is evident from the figures that the nonparametric estimates of the conditional mean approximate the true model in (4.1) extremely well for both n = 50 and 400 but especially for n = 400. The plot for n = 100 was similar to that for n = 400. Though not plotted here, the 95% confidence interval, based on (3.8), contained the

FIG. 1

NONPARAMETRIC REGRESSION (N=50)

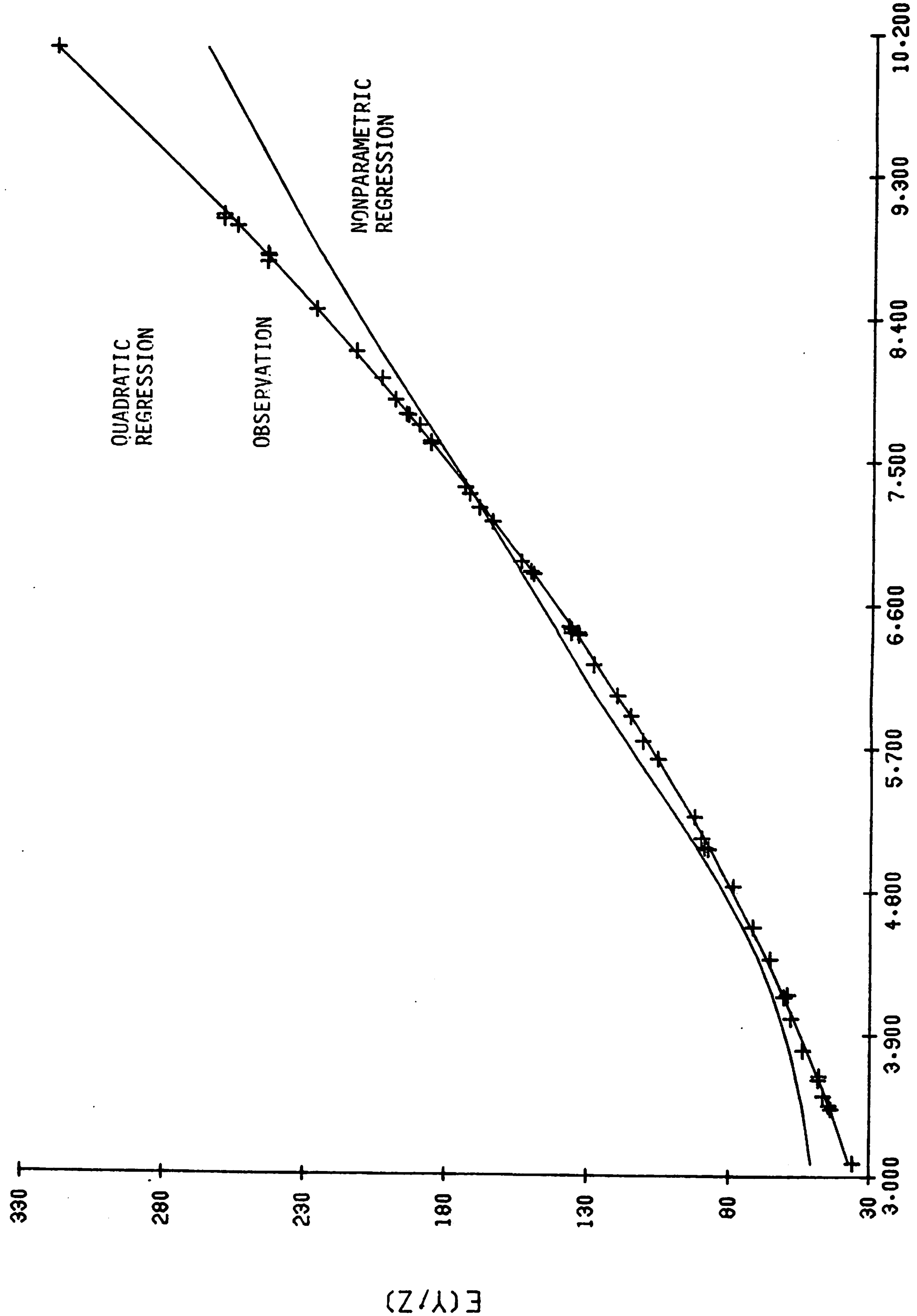
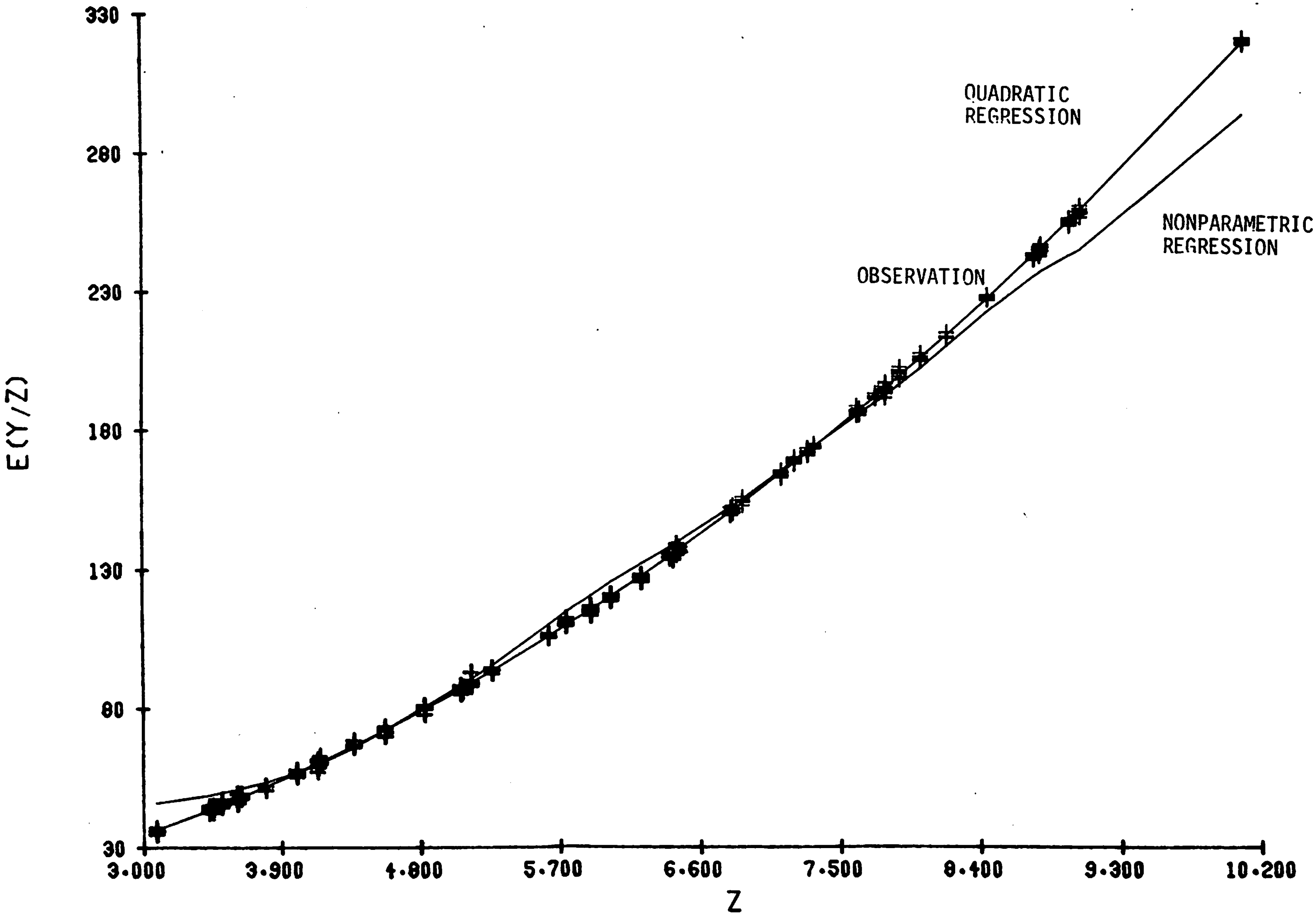


FIG. 2
NONPARAMETRIC REGRESSION (N=400)



true model except in the tail ends. Similar results were obtained with various other choices of K discussed in Section 2.

Figure 3, for n = 400, shows that the nonparametric estimate of the $V(u|_z)$ fluctuates very closely around the true value 1. Thus, again the nonparametric estimate performed well.

The implication of the above results is that the nonparametric estimates can be taken as a reasonably good approximation of the conditional mean (econometric model or the functional form) and variance when in actual practice, as in the econometric issues of the following sections, they are not known.

4.2 Forecast and Variability of Inflation

The questions of accurate forecasting and the variability of inflation rates are of significant interest for the macroeconomists and the policymakers in the government and industries. In fact, Friedman's Nobel Prize Lecture (1977) ascribe real effects in an economy to a higher rate of inflation if that higher rate is accompanied by increased variability.

Both the issues of forecasting and variability can be analyzed by the nonparametric estimates of the conditional mean and variance, that is,

$$\hat{E}(P_t|I_{t-1}) \text{ and } \hat{V}(P_t|I_{t-1})$$

where P_t is the inflation rate and I_{t-1} is the information available up to the period t-1. For this purpose we considered the U.S. decade data (1750-1980) on the wholesale price index (see Batra (1985, p. 85)).

For the purpose of forecasting we estimated $\hat{E}(P_t|P_{t-3})$ by using (3.4) and the kernel function in (4.4). The choice of P_{t-3} was due to the fact that the inflationary peaks appeared roughly after every three decades, see Figure 3. Our calculations provided $\hat{E}(P_{1990}|P_{1960}) = .031$ and $\hat{E}(P_{2000}|P_{1970}) = .029$. These results are consistent with the cycle of inflation in Figure 4 and the conjecture of low inflation rates during 1990-1994 indicated in Batra (1985, p. 95).

The variability in inflation, $\hat{V}(P_t|I_{t-1})$ can be obtained by using (3.17). For this purpose we considered I_{t-1} as P_{t-1} and calculated

FIG. 3

VAR(U/Z) AGAINST Z (N=400)

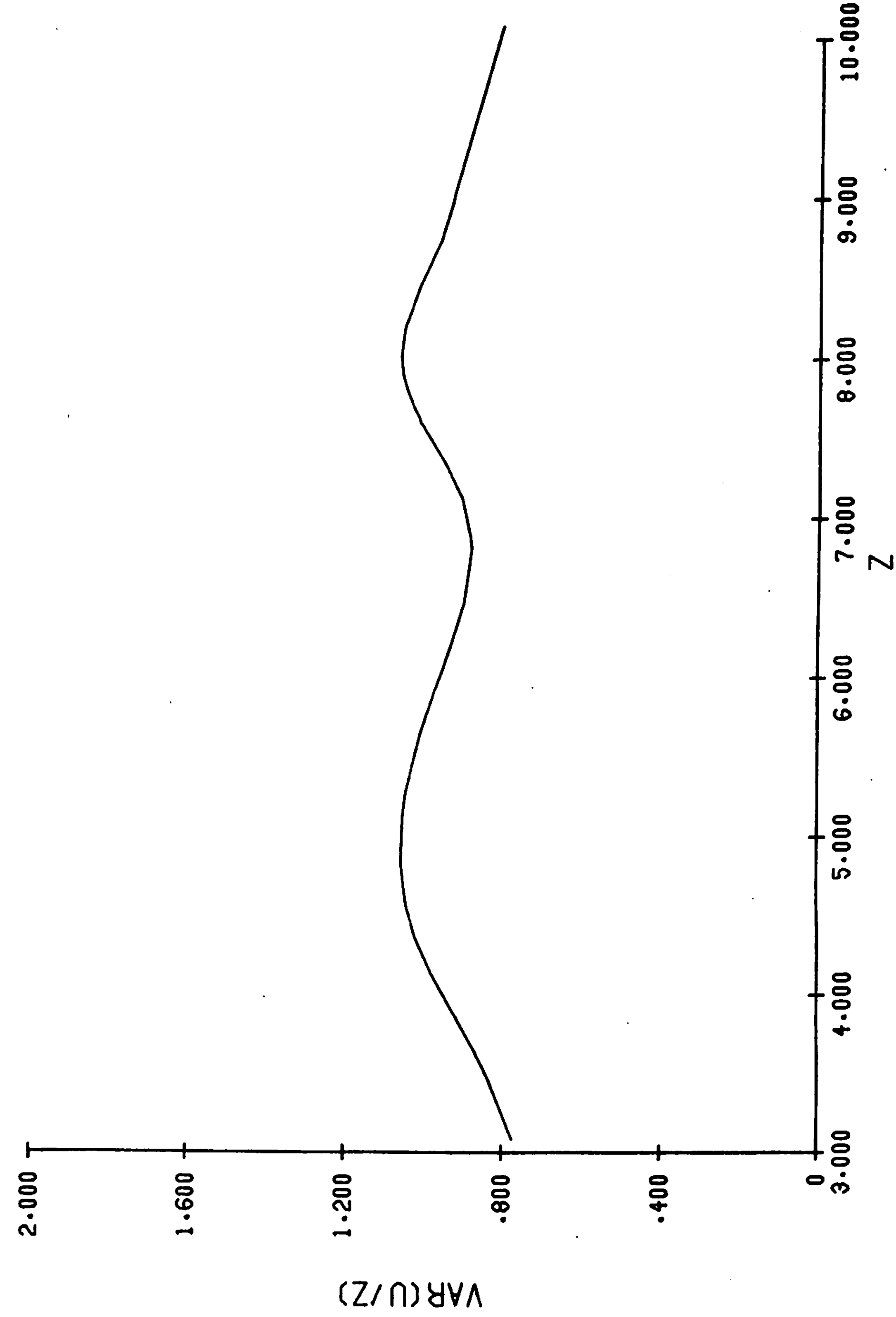
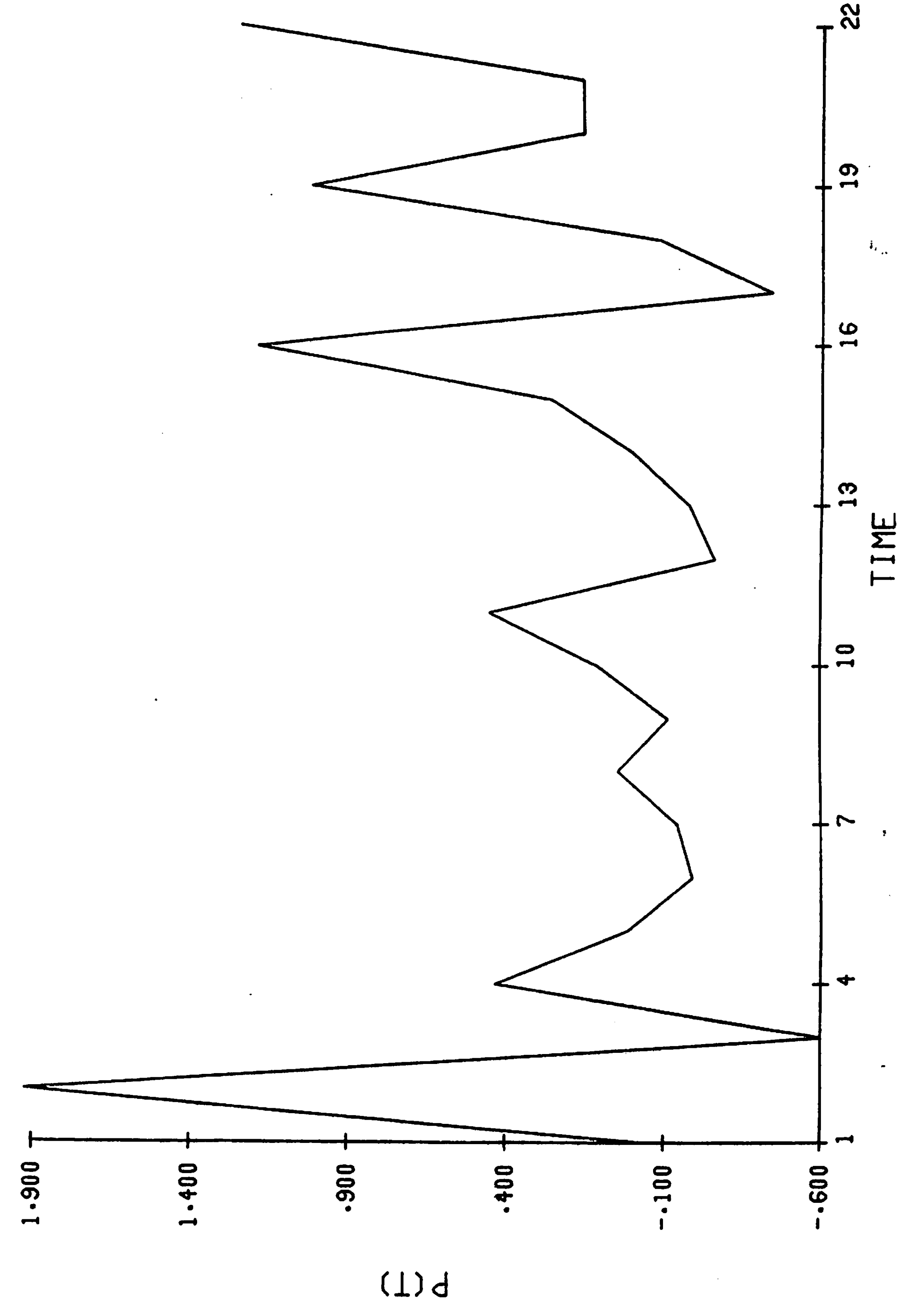


FIG. 4

US INFLATION NONPARAMETRIC REGRESSION



$\hat{V}(P_t|P_{t-1})$ as well as $\hat{E}(P_t|P_{t-1})$. The plots of these values in Figure 5 support the hypothesis that, in the U.S. case, variability does increase with the anticipated inflation. A similar result was also found on the basis of the plot using unanticipated inflation, $\hat{P}_t = P_t - \hat{E}(P_t|_{t-1})$ instead of P_t. The positive relationship between $\hat{V}(\ )$ and $\hat{E}(\ )$ has been referred as absolute variability by Pagan, et al (1983).

4.3 Test for Random Walk

Consider the first-order stochastic difference equation as

$$y_t = \alpha y_{t-1} + u_t, \qquad t=1,\ldots,n \tag{4.5}$$

where u_t are independently and identically $N(0,\sigma^2)$ disturbance terms. When $\alpha = 1$ then y_t is said to follow a random walk. The random walk hypothesis plays an important role in rational expectation hypothesis, and this hypothesis has been investigated in a variety of empirical studies in time series economics. For the model (4.5), the random walk hypothesis $\alpha = 1$ has been investigated recently by Evans and Savin (1981).

Note that the random walk hypothesis $H_o:\alpha = 1$ implies

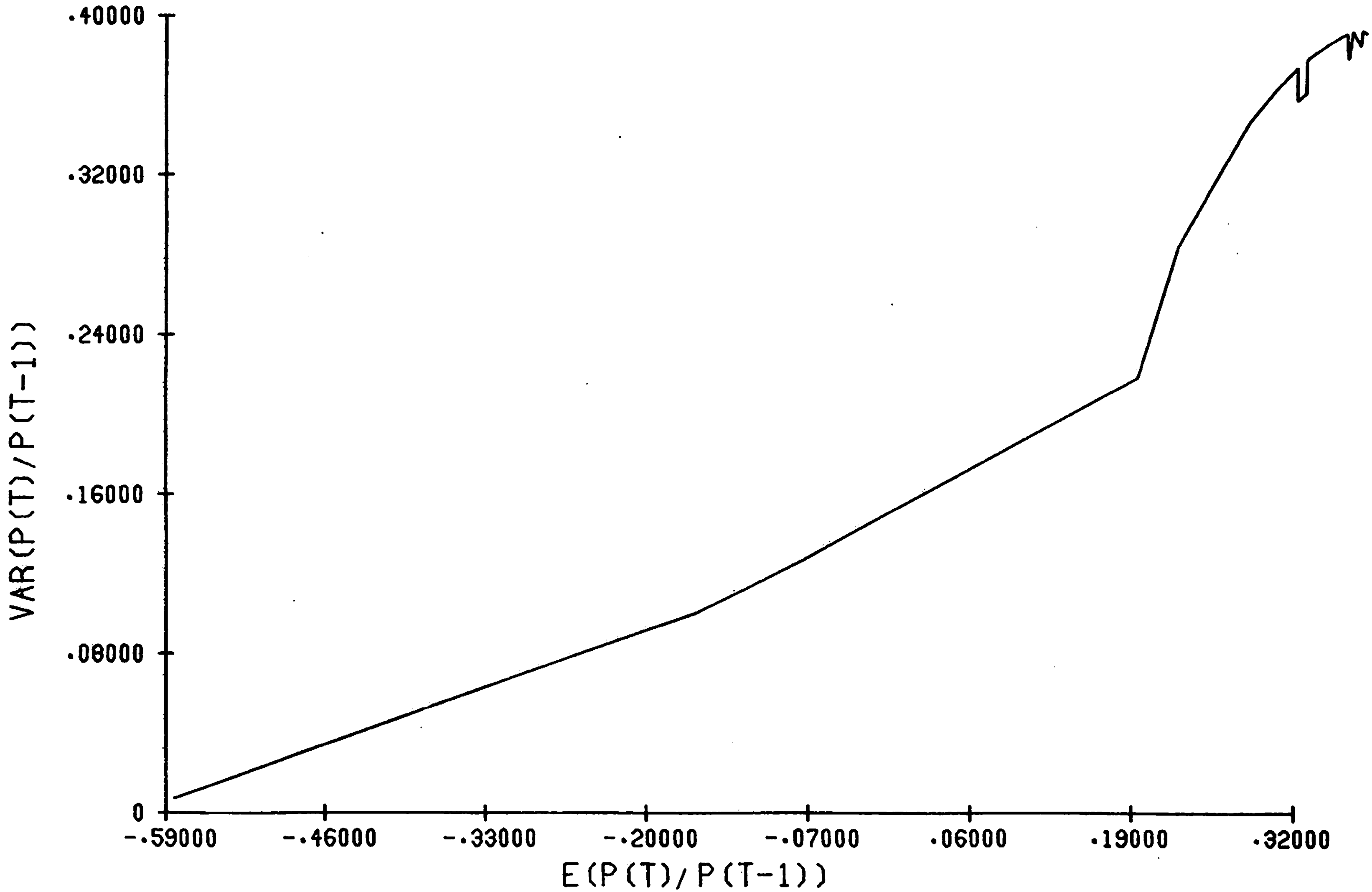
$$H_o: E(y_t|y_{t-1}) = y_{t-1} .$$

This hypothesis can therefore be analyzed by obtaining the nonparametric estimate of $E(y_t|y_{t-1})$ and calculating the confidence interval for $E(y_t|y_{t-1}) = y_{t-1}$ given in (3.9).

For an illustration we analyzed the random walk hypothesis for the Canadian-U.S. monthly spot rates data that were kindly supplied by the Bank of Canada, and used in Longworth (1981). Considering y_t= log of spot rate = log s_t we estimated $E(y_t|y_{t-1})$, using kernel in (4.4), and its 95% C.I. Since under $H_o: Ey_t|y_{t-1} = y_{t-1}$, Figure 6 does not support the random walk hypothesis. This is because most of the y_{t-1} points are outside the confidence bands. The implication of this result is that the past spot rate is not necessarily a good predictor of the current spot rate.

FIG. 5

US INFLATION NONPARAMETRIC REGRESSION



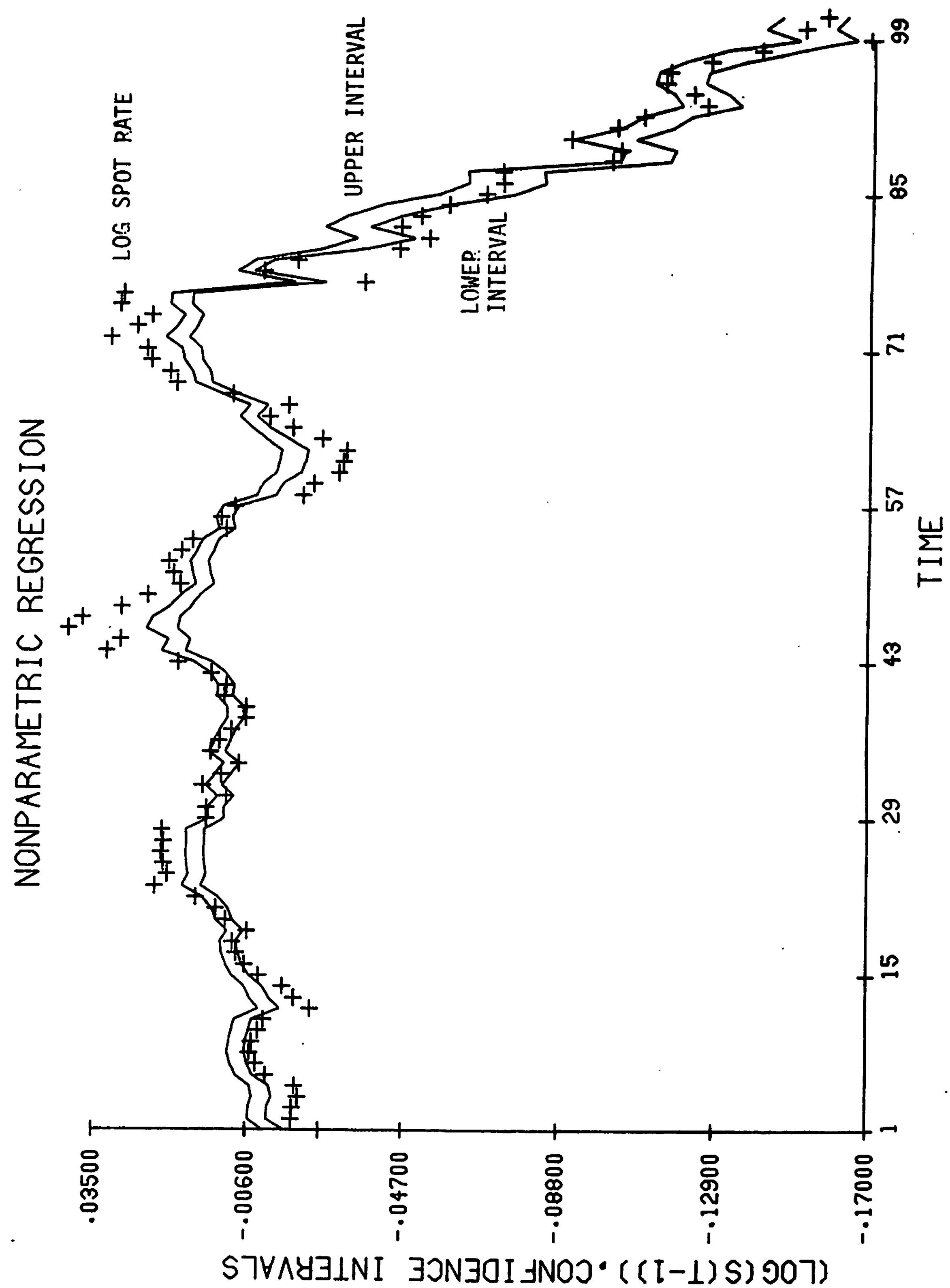

NONPARAMETRIC REGRESSION
LOG SPOT RATE
UPPER INTERVAL
LOWER INTERVAL
TIME
(LOG(S(T-1)).CONFIDENCE INTERVALS
.03500
-.00600
-.04700
-.08800
-.12900
-.17000
1
15
29
43
57
71
85
99

FIG. 6

5. CONCLUSION

We have reviewed the literature on the class of nonparametric density estimators. The application of these estimators in the estimation of regression functions, model specification, specification testing, and finite sample econometrics has been explored. Illustrative examples of the estimation of regression function, variability, forecasting and random walk hypotheses are presented.

The list of applications, in addition to ones provided in Section 3, are large and all of them cannot be presented here. In fact one can analyze any econometric issue by nonparametric methods, and lead to better inference for policy purposes. Some of the other areas of applications which can be mentioned are the estimation of quantile functions (see Parzen (1979)), estimation of hazard functions (Tanner and Wong (1984), Heckman and Singer (1984)), estimation of entropy measures, and estimation of the Fisher information matrix (Singh (1977)). The future areas of applications are the estimation of rational expectations models and nonlinear simultaneous equations models.

ACKNOWLEDGMENTS

The authors are grateful to the NSERC for research support, and to Virginia Ho for excellent research assistance. They are also thankful to R. Carter, D. Poirier and J. Mackinnon for useful comments on an earlier version of this paper.

BIBLIOGRAPHY

Amemiya, T. (1973), "Regression Analysis When the Dependent Variable is Truncated Normal," Econometrica 41, 997-1016.

Bartlett, M. S. (1963), "Statistical Estimation of Density Function," Sankhya Ser. A 25, 245-254.

Batra, R. N. (1985), Regular Cycles of Money, Inflation, Regulation and Depressions, Venus Books, Dallas.

Bhattacharya, P. K. (1967), "Estimation of a Probability Density Function," Sankhya Ser. A 29, 373-382.

Burbea, J. and C. R. Rao (1982), "Entropy Differential Metric, Distance and Divergence Measures in Probability Spaces: A Unified Approach," J. Multivariate Anal., 575-596.

Cacoullos, T. (1966), "Estimation of a Multivariate Density," Ann. Inst. Statist. Math. 18, 176-189.

Carroll, R. J. (1976), "On Sequential Density Estimation," Z. Wahrschein--Lichkeitstheorie und Verw. Gebiete 36, 137-151.

Cohen, A. C. Jr. (1950), "Estimating the Mean and Variance of Normal Populations from Singly and Doubly Truncated Samples," Ann. Math. Statist., 557-569.

Cosslett, S. (1978), "Efficient Estimation of Discrete-Choice Models from Choice-Based Samples," Ph.D. Dissertation, Department of Economics, University of California - Berkeley.

_______ (1980), "Estimation of Random Utility Models with Unknown Probability Distribution," Paper presented to CEME Conference on Discrete Econometrics.

Davidson, R. and J. G. Mackinnon (1981), "Several Tests for Model Specification in the Presence of Alternative Hypothesis," Econometrica 49, 781-1244.

Davies, H. L. (1973), "Strong Consistency of a Sequential Estimator of a Probability Density Function," Bull. Math. Statist. 15, 49, 54.

Davies, H. L. and E. J. Wegman (1975), "Sequential Nonparametric Density Estimation," IEEE Trans. Info. Theory IT-21, 619-628.

Deheuvels, P. (1974a), "Conditions Necessaries et Suffisantes de Convergence Ponctuelle Presque Sure et Uniforme Presque Sure des Estimateurs de la Densite," C.R. Acad. Paris Ser. A 278, 1217-1220.

_______ (1974b), "Estimation Sequentielle de la Densite," Universite de Paris VI.

Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," Ann. Statist. 7, 1-26.

Engle, R. F., D. F. Hendry and J. F. Richard (1983), "Exogeneity," Econometrica 51, 277-304.

Epanechnikov, V. A. (1969), "Nonparametric Estimates of a Multivariate Probability Density," Theor. Probab. Appl. 14, 153-158.

Evans, G. B. A. and N. E. Savin (1981), "Testing for Unit Roots: 1," Econometrica 49, 753-779.

Fair, R. C. (1973), "A Comparison of FIML and Robust Estimates of a Nonlinear Econometric Model," National Bureau of Economic Research Working Paper No. 15, October.

Friedman, M. (1977), "Nobel Lecture: Inflation and Unemployment," J. Political Economy 85, 451-472.

Hausman, J. (1978), "Specification Tests in Econometrics," Econometrica 46, 1251-2171.

Hausman, J. A. and D. A. Wise (1976), "Social Experimentation, Truncated Disturbances and Efficient Estimation," Econometrica 45, 319-339.

_______ (1978), "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," Econometrics 46, 403-426.

Heckman, J. (1976), "Simultaneous Equations Model with Continuous and Discrete Endogenous Variables and Structural Shifts," in Goldfeld and Quant (eds.), Studies in Non-Linear Estimation, Cambridge: Ballinger.

Heckman, J. and B. Singer (1984), "Econometric Duration Analysis," J. Econometrics 24, 63-132.

Johns, M. F. and J. Van Ryzin (1972), "Convergence Rates for Empirical Bayes Two-Action Problems II. Continuous Case," Ann. Math. Statist. 43, 934-947.

Kronmal, P. and M. Tarter (1968), "The Estimation of Probability Densities and Cumulatives by Fourier Series Methods," J. American Statist. Assoc. 38, 482-493.

Lee, L. F. (1982), "Specification Error in Multinormal Logit Models: Analysis of the Omitted Variable Bias," Jl Econometrics 20, 197-210.

Loftsgaarden, D. O. and Quesenbury, C. P. (1965), "A Nonparametric Estimate of a Multivariate Density Function," Ann. Math. Statist. 36, 1049-1051.

McFadden, D. (1984), "Econometric Analysis of Qualitative Response Models," Advances in Econometrics, Vol. II, Cambridge University Press.

Moore, D. S. and J. W. Yackel (1977), "Consistency Properties of Nearest Neighbor Density Estimates," Ann. Statist. 5, 143-154.

Nadaraya, E. A. (1965), "On Nonparametric Estimates of Density Function and Regression Curves," Theor. Probab. Appl. 10, 186-190.

Pagan, A. R., A. D. Hall and P. K. Trivedi (1983), "Assessing the Variability of Inflation," Rev. Economic Stud.

Parzen, E. (1962), "On the Estimation of Probability Density and Model," Ann. Math. Statist. 33, 1065-1076.

_______ (1979), "Nonparametric Statistical Data Modeling," Journal of the Amer. Statist. Assoc. 74, 105-131.

Phillips, P. C. B. (1983), "Exact Small Sample Theory in the Simultaneous Equations Model," Advances in Econometrics, Vol. 1, Cambridge University Press.

Rao, B. L. S. P. (1969), "Estimation of Unimodal Density," Sankhya 31, 23-36.

Robinson, P. M. (1983), "Nonparametric Estimators for Time Series," J. Time Ser. Anal. 85-208.

Rosenblatt, M. (1956), "Remarks on Some Nonparametric Estimates of Density Function," Ann. Math. Statist. 27, 832-837.

Schucany, W. R., H. C. Gray and D. B. Owen (1971), "On Bias Reduction in Estimation," J. Amer. Statist. Assoc. 66, 524-533.

Schucany, W. R. and J. P. Sommers (1976), "Improvement of Kernel Type Density Estimators," J. Amer. Statist. Assoc. 71, 420-433.

Schuster, E. F. (1969), "Estimation of a Probability Density Function and its Derivatives," Ann. Math. Statist. 40, 1187-1195.

Schwartz, S. C. (1967), "Estimation of a Probability Density by an Orthogonal Series," Ann. Math. Statist. 38, 1261-1265.

Singh, R. S. (1974), "Estimation of Derivatives of Average of μ-Densities and Sequence-Compound Estimation in Exponential Families," RM-318, Dept. of Statistics and Probability, Michigan State University, East Lansing, MI, U.S.A.

_______ (1977a), "Improvement on Some Known Nonparametric Uniformly Consistent Estimators of Derivatives of a Density," Ann. Statist. 5, 394-399.

_______ (1977b), "Applications of Estimators of a Density and its Derivatives to Certain Statistical Problems," J. Roy. Statist. Soc. Ser. B 39, 357-363.

_______ (1978), "Nonparametric Estimation of Derivatives of Average of μ-Densities with Convergence Rates and Applications," SIAM J. Appl. Math. 35, 637-649.

_______ (1979a), "Empirical Bayes Squared Error Loss Estimation in Exponential Families and the Best Possible Rates," Ann. Statist. 7, 890-902.

Singh, R. S. (1979b), "Mean Squared Errors of Estimates of a Density and its Derivatives," Biometrika 66, 177-180.

_______ (1980), "Asymptotic Normality of Nonparametric Estimators of Derivatives of Average of μ-Densities," SIAM J. Appl. Math. 39, 431-439.

_______ (1981), "Speed of Convergence in Nonparametric Estimation of a Multivariate μ-Density and its Mixed Partial Derivatives," J. Statist. Planning Inf. 5, 287-298.

Singh, R. S. and A. Ullah (1984), "Estimation of Multivariate DGP," University of Western Ontario Manuscript.

_______ (1985), "Nonparametric Time Series Estimation of Joint DGP, and Vector Autoregression," Econometric Theory 1.

Susarla, V. and S. Kuwar (1975), "Estimation of Partial Derivatives of the Average of Densities," J. Austral. Math. Soc. 20, 230-241.

Tanner, M. A. and W. H. Wong (1984), "Data-Based Nonparametric Estimation of the Hazard Function with Applications to Model Diagnostic and Exploratory Analysis," J. Amer. Statist. Assoc. 79, 174-182.

Tobin, J. (1958), "Estimation of Relationships for Limited Dependent Variables," Econometrica 26, 24-36.

Ullah, A. (1983), "Divergence, Distance and Entropy Measures: Unification and Applications," University of Western Ontario Manuscript.

_______ (1985), "Specification Analysis of Econometric Models," J. Quantitative Econ. forthcoming.

Van Ryzin, J. (1973), "On Histogram Method of Density Estimation," Comm. Statist. 2, 493-506.

Watson, G. S. (1955), "Serial Correlation in Regression with Dependent Observations," Econometrica 52, 143-161.

Wegman, E. J. (1970), "Maximum Likelihood Estimation of Unimodal Density Function," Ann. Math. Statist. 41, 457-571.

Whittle, P. (1958), "On Smoothing of Probability Density Functions," J. Roy. Statist. Soc. B 20, 334-343.

Yamato, H. (1971), "Sequential Estimation of a Continuous Probability Function and Mode," Bull. Math. Statist. 14, 1-12.